Smart Agricultural System

A Lightweight AI-on-Device Solution for Smallholder Farmers

Authors

Yixin Yang (122090900)
Xinran Deng (121020038)
Minghao Liu (121090344)
Yike Zhang (122030105)

May 11, 2025

Acknowledgement:

We are grateful to our supervisor, Professor *Guo Yutong*.

Contents

Executive Summary

Smallholder farmers are the backbone of Chinese agriculture, managing over 70% of the nation's farmland and comprising 90% of its agricultural workforce (Wang, 2023). Yet they face critical challenges: aging farm populations—the average age of a Chinese farmer is over 53—with limited access to credit and technology, persistent crop pests and diseases, and pressure to improve yields sustainably (Donnellon-May, 2025).

This report proposes a "Smart Agricultural System" – a lightweight, AI-powered pest and disease detection solution – to address these challenges in a practical, commercially viable way. The system centers on a MobileNetV3-based Convolutional Neural Network (CNN) model, optimized and quantized to run entirely on affordable devices (e.g. smartphones) offline, without need for internet connectivity.

This on-device approach enables real-time crop monitoring in remote fields, reducing reliance on costly infrastructure or external services. While the system is designed primarily to empower individual smallholder farmers with actionable, expert-level diagnostics at their fingertips, it also serves as a modular tool that agricultural technology enterprises can embed in autonomous field patrol units, aerial drones, or smart spraying platforms, enabling distributed edge inference and multi-device collaboration.

The report is structured as a comprehensive 12-week business-AI course project. It begins by analyzing the context of Chinese smallholder farming and the role of AI in modern agriculture. We review existing AI agriculture solutions (from simple smartphone apps to advanced drone and IoT systems) and identify their limitations for smallholders—often high cost, complexity, or connectivity needs.

We then detail the proposed solution's technical design: a MobileNetV3 CNN trained to recognize crop pests and diseases from leaf images, compressed via quantization for fast edge inference. Key benefits of this solution include early pest/disease detection (reducing crop losses), targeted intervention (minimizing pesticide overuse), and empowerment of farmers with actionable data (FAO, 2020).

We further assess business feasibility, showing strong market need and alignment with China's agricultural modernization goals. A roadmap is outlined for implementation: data collection, model development, pilot deployment, and scaling. Risk analysis covers model accuracy, user adoption, and ethical considerations (such as data privacy and ensuring equitable access).

In conclusion, the report highlights the system's potential to improve smallholder livelihoods and productivity, and discusses future extensions (integration with IoT sensors and weather forecasts, broader crop coverage) to sustain impact. By focusing on a lightweight, user-friendly AI solution, this project illustrates how advanced AI can be applied pragmatically to real-world farming challenges, delivering social and economic value in China's rural heartland.

In conclusion, the report highlights the system's potential to improve smallholder livelihoods and productivity, and discusses future extensions (integration with IoT sensors and weather forecasts, broader crop coverage) to sustain impact. By focusing on a lightweight, user-friendly AI solution, this project illustrates how advanced AI can be applied pragmatically to real-world farming challenges, delivering social and economic value in China's rural heartland.

1 Introduction

1.1 Challenges in Chinese Agriculture

China is a nation of numerous smallholder farmers, with around 200–300 million farming households cultivating only a few hectares each (Wang, 2023). These smallholders are essential to national food security, yet they face persistent constraints.

Demographic shifts have contributed to a rapidly aging rural workforce—the average Chinese farmer is over 53 years old, with more than a quarter aged above 60 (Donnellon-May, 2025). This raises long-term concerns about the viability of labor-intensive farming. At the same time, many smallholders operate under financial pressure. Nearly 19% of family farms report operational funding gaps (Wang, 2023), limiting investment in quality inputs or modern technologies.

Another serious issue is the growing frequency of pest and disease outbreaks, intensified by climate change. Globally, up to 40% of crop output is lost annually due to such biotic stresses (FAO, 2020). In response, farmers in China often rely on broad-spectrum pesticide use. However, this approach has diminishing returns and harmful side effects, including environmental degradation and threats to human and ecological health.

Compounding these challenges is the limited access to agronomic knowledge and support services in rural areas. With extension systems overstretched, many farmers rely on experience or guesswork, leading to suboptimal decisions regarding pest control, irrigation, and harvest timing. Together, these factors expose the structural vulnerabilities of smallholder agriculture and highlight the need for smarter, more accessible support tools.

1.2 Rationale for AI Application

Artificial intelligence offers a promising path to address the scale, complexity, and urgency of smallholder farming challenges. Among recent innovations, image-based AI models—such as convolutional neural networks (CNNs)—have proven especially effective at detecting plant diseases from leaf photos, enabling early and accurate intervention.

What makes AI particularly compelling in this context is its scalability and accessibility. Lightweight, quantized models can now be deployed directly on smartphones or low-power devices, eliminating the need for high-speed internet or external experts. This makes AI uniquely positioned to bridge the knowledge and resource gap that smallholders face.

Beyond technical feasibility, AI also supports broader sustainability goals. By improving targeting and diagnosis, it reduces unnecessary pesticide use and promotes data-driven decisions. For China's millions of dispersed smallholder farms, AI provides not just automation, but accessible expertise—on demand, at the edge, and at scale.

2 Review of Existing AI Solutions and Their Limitations

AI-driven agriculture solutions range from basic smartphone applications to complex systems involving drones and edge computing devices. It is important to assess these existing solutions and why many have not yet achieved widespread adoption among smallholder farmers in China. Below, we examine a few representative categories of AI solutions for pest and disease management, and identify their key limitations.

1. Mobile Apps with Cloud AI Services: Several mobile applications allow farmers to take a photo of a diseased plant and upload it to a cloud server where an AI model diagnoses the issue. One of the example is Plantix, an application you can find on Google Play, which leverage powerful cloud-based neural networks to recognize plant diseases with high accuracy(Plantix, 2023). Although this approach can yield accurate results, it has practical drawbacks for rural farmers. It requires a reliable internet connection to send images and receive diagnoses – a challenge in remote farming villages with limited connectivity. The turnaround time may also be slow if network coverage is poor. Moreover, dependence on cloud services introduces ongoing costs (subscription or data fees) that cash-strapped smallholders may not afford. There are also concerns about data privacy and sovereignty when farmers' field data is stored on external servers. The result is that purely cloud-based solutions, despite their technical merits, see limited uptake in regions without strong digital infrastructure.

2. Drone Imaging Solutions: Drones equipped with AI-powered cameras can survey large areas quickly, identifying pest infestations or water stress from the air. In China, some larger farms use drone patrols to monitor crops, which can then trigger precision spraying. Although the drone effective for large-scale precision agriculture, this solution is less accessible to an individual smallholder. Deploying drones is very costly and require skilled operation. Smallholders with landholdings often under 1 hectare may not see sufficient return on investment. Additionally, interpreting drone data often needs expert analysis or advanced software. Thus, while drone-based AI monitoring is promising, for many small farmers it remains out of reach without government programs or cooperatives to provide shared access. It also typically functions as part of a broader system (with cloud analytics), again raising the connectivity issue.

3. IoT Sensor Networks with AI: Another approach is deploying IoT devices (like smart camera traps or sensor nodes) in the field that use on-device AI to detect pests. One example is the "smart trap" for fruit orchards developed by researchers, which uses a low-power camera and an on- board neural network to identify insects lured to a trap. All computation is done on the node itself, drastically reducing data transmission; only a small notification is sent when a pest is detected (Albanese, A., Nardello, M., Brunelli, D., 2021) . These edge devices often include solar panels and battery to operate autonomously in rural area.

The advantage is clear: continuous, automated monitoring without needing internet or constant human oversight. In tests, such systems have shown they can run indefinitely with energy harvesting and detect infestations in real-time. However, current implementations are mostly experimental or in pilot stages. The limitations include the complexity of deployment (installing and maintaining devices across many small farms) and cost – a Raspberry Pi-based trap with neural accelerator and solar power, as in the research prototype, though cheaper than constant manual scouting, is still a significant investment for an individual farmer. Additionally, these devices may be designed for specific pests (e.g., moths in pheromone traps) and not easily adaptable by farmers for different crops or diseases without technical support. Thus, while AI traps solve connectivity and accuracy issues, they are not yet plug-and-play products for the typical smallholder.

Comparative Limitations: Figure 1 summarizes how existing approaches stack up for smallholder farmers.

Solution	Pros	Cons / Limitations
Manual monitoring (traditional)	No special equipment needed; farmer' s local knowledge	Labor-intensive; often detects problems late; less accurate; high crop loss and chemical use.
Cloud-based AI apps	High accuracy AI models; user-friendly if internet is available	Requires internet/smartphone; potential latency; data costs; farmers reliant on external service; privacy concerns.
Drone surveys	Covers large areas quickly; can integrate with precision sprayers	High cost equipment; requires expertise or service provider; not practical for very small plots; data interpretation needed.
AI-powered IoT devices (edge traps)	Real-time autonomous monitoring; works offline; low data transmission; very precise targeting	Initial setup cost; limited to specific pests or use-cases; maintenance and technical know-how required; currently in pilot stage.

Figure 1: Overview of existing AI solutions for agriculture.

3 Proposed AI Solutions

Our proposed solution is a Hybrid AI system for pest and disease detection. We do the consumer segmentation first. We hope to provide different services to different types of customers. For those small farmers, we provide an on-device AI system. The core of the system is a MobileNetV3-based CNN model running directly on the user's device (especially for a low-cost edge device). By performing inference on-device, the system works without internet connectivity, Faster response time while protecting user privacy and providing instant results in the field. They can also choose to upload complex problems to the cloud for processing.

For large agricultural companies, we offer a cloud-based AI system, which focus on edge inference and multi device collaboration, and scalable. It provides autonomous field patrol or aerial monitoring, GPS integration for disease localization and compatible with pesticide spraying systems. Drones can be operated at the terminal for pesticide spraying.

4 Technical Components Deep Dive

4.1 CNN + Fine-Tuning

In this project, we primarily adopt a Convolutional Neural Network (CNN) as the core model for image recognition. CNNs are capable of automatically learning local features such as edges, textures, and shapes from images, making them suited for identifying and classifying agricultural pests and diseases. However, since agricultural pest datasets tend to be small and imbalanced, training a CNN from scratch is time-consuming and computationally inefficient. Therefore, we choose to use a fine-tuning strategy. Specifically, we will build upon a CNN that has been pre-trained on a large-scale general image dataset (ImageNet) and further train it on our agricultural pest and disease images. This approach allows us to retain the model's feature extraction capabilities while improving its accuracy for agricultural tasks. As a result, the model will maintain strong generalization performance even with limited training samples.

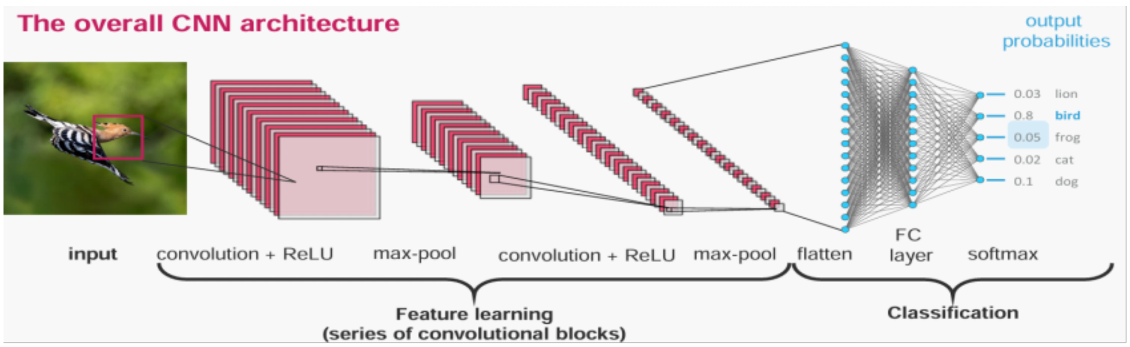

Figure 2: The overall CNN architecture.

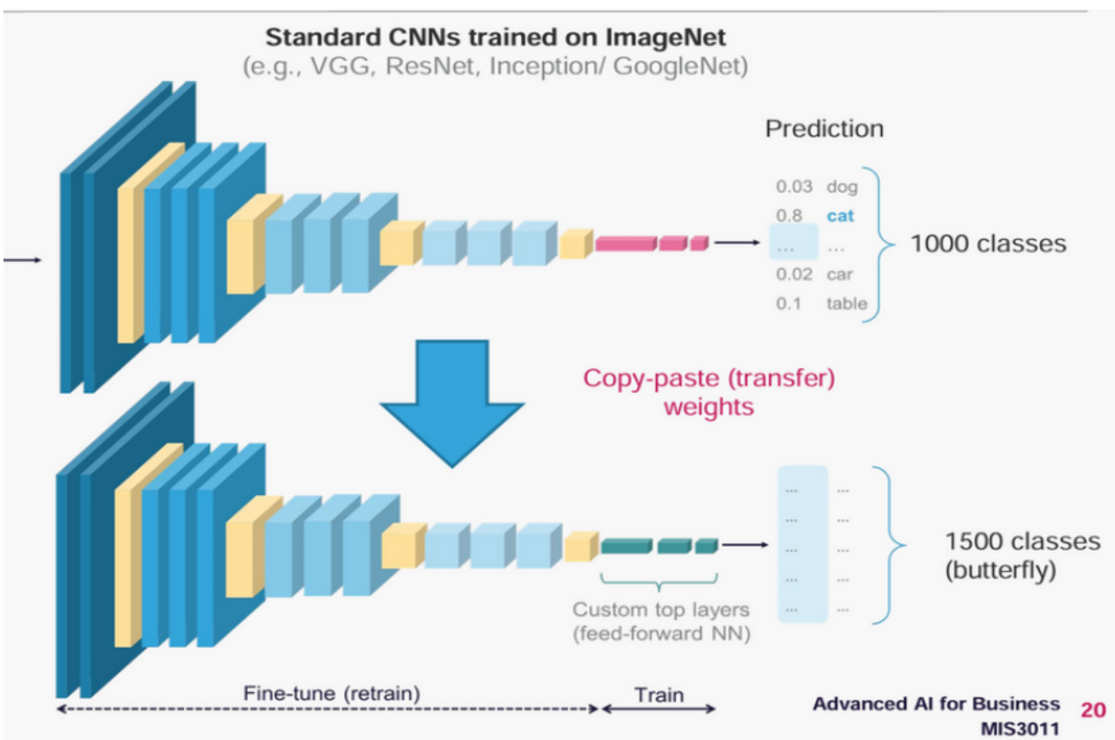

Figure 3: Transfer learning using ImageNet-trained CNNs.

4.2 MobileNetV3

MobileNetV3 is a lightweight neural network model specifically designed for mobile and edge devices. It features three key components. First, it replaces the traditional ReLU activation function with the h-swish activation function, which enhances model performance without increasing computational cost. In addition, MobileNetV3 introduces the SE (Squeeze-and-Excitation) attention module, which automatically learns the importance of each channel and assigns higher weights to critical features, thereby improving the model's attention mechanism. Most importantly, the model has a quantization- and pruning-friendly architecture. Its structure is naturally compatible with subsequent model compression and acceleration operations, making it suited for INT8 quantization and redundant channel pruning. This significantly reduces model size and speeds up inference, paving the way for efficient local deployment.

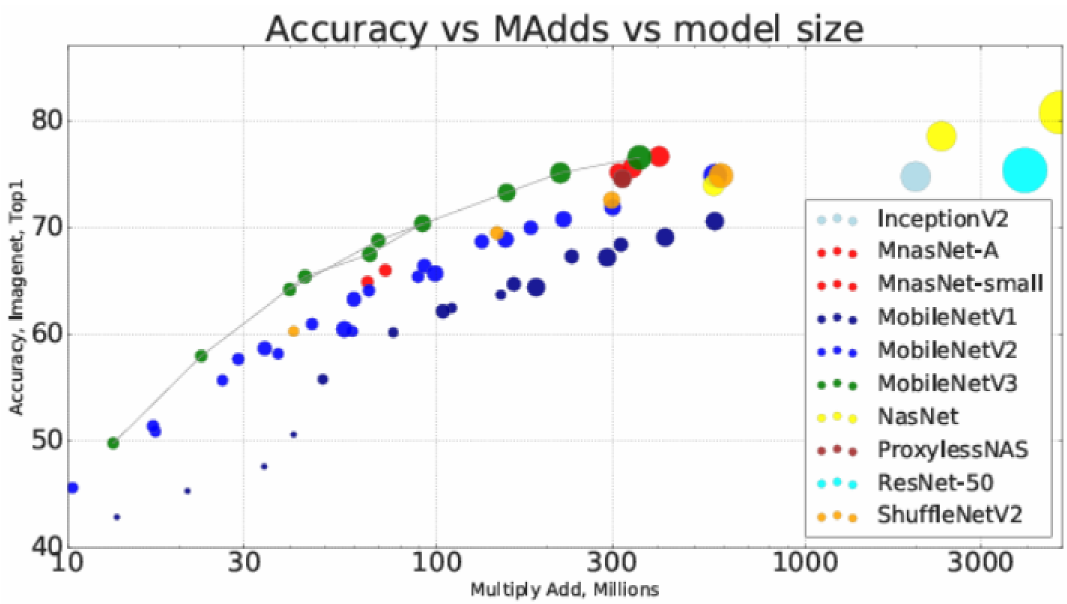

Figure 4: Accuracy vs. MACs vs. model size for mobile models.

4.3 Compression and Optimization

A crucial step for on-device deployment is model compression the model. Our goal is to minimize our model to be small enough (ideally a few megabytes) to run inference in under a second on a typical smartphone. The first compression method we will apply is quantization. Quantization means reducing numerical precision from floating point to integers, decreasing model size while maintaining accuracy. We will use 8-bit quantization on the trained MobileNetV3 model, converting it from 32-bit floating point to 8-bit integer representations. Quantization dramatically reduces the model size and allows it to run efficiently on mobile CPUs that have 8-bit arithmetic accelerators. Modern AI frameworks like TensorFlow Lite or PyTorch Mobile provide quantization tools that can do this with minimal loss in accuracy. Figure 5. The significant improvement from quantization:Figure 6

Another method we will consider is pruning techniques, which removing redundant neurons/weights as well as reducing computational demands. A research has demonstrated that pruning during training can maintain accuracy while reducing complexity, thereby speeding up inference.(Albanese, A., Nardello, M., Brunelli, D., 2021) By iteratively pruning and fine-tuning, the model's computational demand and latency can be minimized. Figure 7 (soucres:https://towardsdatascience.com/pruning-neural-networks-1bb3ab5791f9/)

There are also some potential ways to significantly reduce the size of the model without affecting the accuracy of the model, such as using matrix factorization and decomposing a large matrix into the product of two matrices. The results so far are very encouraging, for example, inference with MobileNetV2 on a Raspberry Pi 3 was feasible, and with hardware acceleration it became even faster(Albanese, A., Nardello,

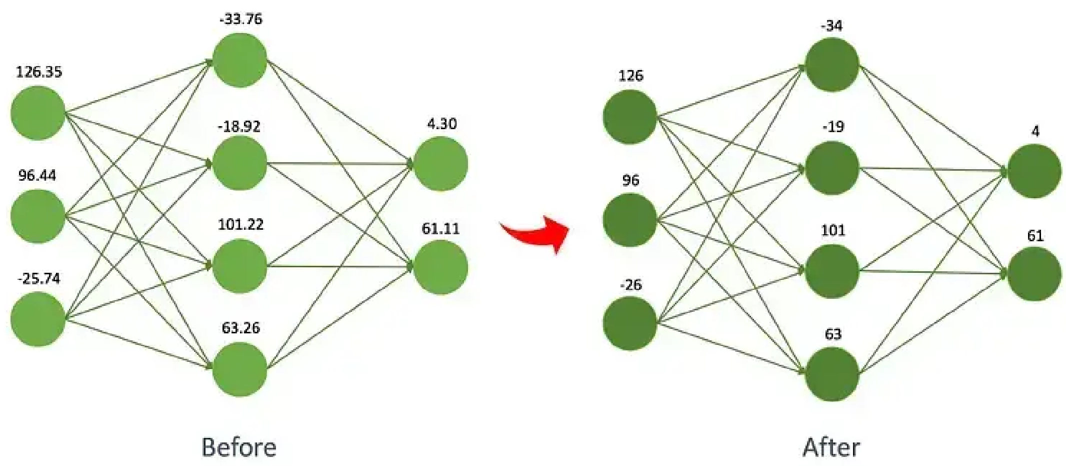

Figure 5: llustration of neural network quantization.

M., Brunelli, D., 2021). MobileNetV3, being even more optimized, should comfortably achieve real-time inference on mid-range mobile devices.

4.4 RAG (Retrieval-Augmented Generation)

After built the LLM model, we still need techniques like RAG (Retrieval Augmented Generation) to solve the adaptation problems. China has a vast territory, and different regions have different crops and pests. To make our AI perform better in different regions, we need to access external databases by RAG. We will access to local meteorological observation websites and local pest and disease related websites, etc. by using API connect.The detailed data strategy will be conclude in next part, so I won't describe it here. Then we cut data into chunks and use word2vec to change the original database into embedding vectors, which our MobileNetV3 model can understand well. When user send a query, we also change it into a vector and search the most similar vectors in our stored vector base.

By applying RAG, we can increase the recall rate and accuracy and decrease the hallucinations problems. Research on integrating AI detection models with language models for real-time pest management in tomato cultivation highlighted the effectiveness of combining YOLOv8 with language models. The study reported precision and recall rates exceeding 98% for detection tasks.(Dr. Selva Kumar S Imadh Ajaz Banday Vibha Venkatesh Shanbhag, 2024) And this research also found that the integration of RAG with YOLOv8 for disease identification significantly reduced hallucinations in LLMs. Figure 8

In summary, the proposed technical solution marries cutting-edge AI (MobileNetV3 CNN) with an edge computing approach (on-device inference) to create a tool that is accurate, fast, and usable in the real- world context of Chinese smallholder farming. By carefully fine-tuning the model, employing transfer learning, optimizing via quantization, as well as applying RAG technique, we ensure the AI is both powerful and lightweight. The next sections will shift focus to the data strategy and implementation aspects – how we prepare the data and how this solution can be delivered and scaled to truly benefit the farmers on the ground.

5 Data Strategy

5.1 Data Sources

We will use an agricultural pest and disease image dataset as our core data source. Specifically, we will obtain crop and pest images from the Agricultural Disease and Pest Research Gallery provided by the Chinese Academy of Sciences' Image Understanding and Machine Learning Database.

($http://www.icgroupcas.cn/website_bchtk/chazhao.aspx$)

Additionally, by leveraging Retrieval-Augmented Generation (RAG) techniques, we aim to integrate local agricultural station data to capture region-specific pest and disease patterns and enhance localization.

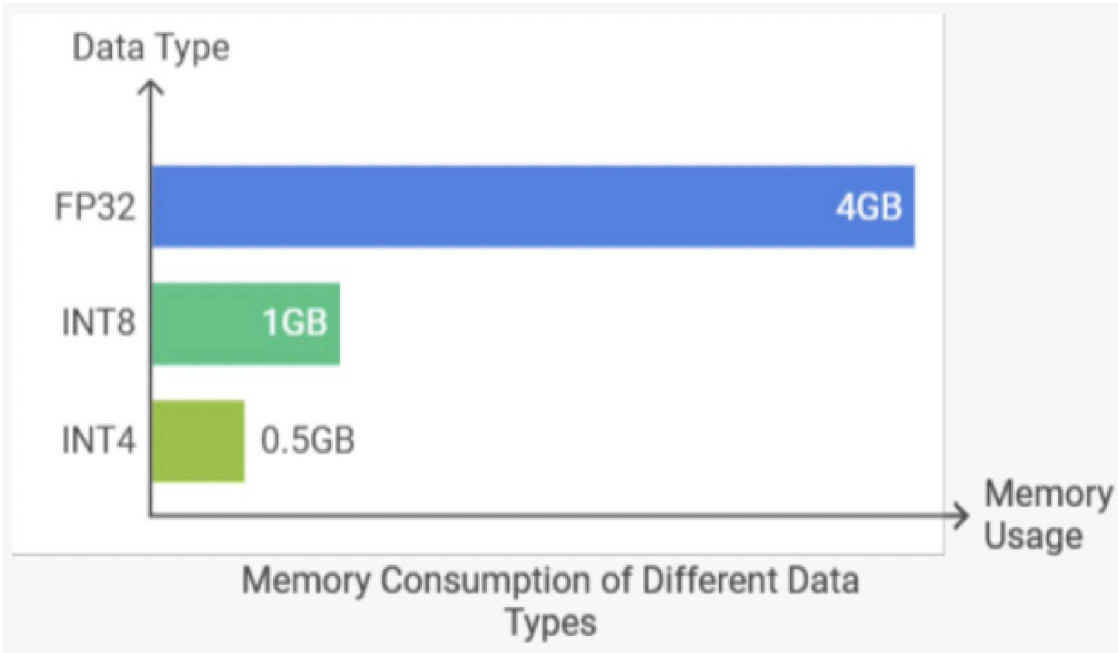

Figure 6: Memory consumption of different numerical formats.

Farmers can also contribute by uploading images of newly observed crop conditions. These real-time image contributions will help keep our dataset continuously updated. Furthermore, we plan to collect historical field monitoring records and weather data from the region to incorporate spatiotemporal context and enhance predictive modeling.

5.2 Data Pre-processing and Annotation

After collecting raw images, we will apply standardized preprocessing steps. First, all images will be resized to a fixed input size (224×224 pixels) to match the input requirements of our MobileNetV3-based model. Next, data augmentation will be applied, including brightness adjustment, contrast enhancement, rotation, and horizontal flipping. These augmentations aim to improve the model's robustness under varying lighting conditions.

We will also filter out noisy or unusable images by applying thresholds on brightness and blur metrics (e.g., using Laplacian variance). Afterward, RGB normalization will be performed to stabilize model training.

Once preprocessing is complete, we will proceed to annotation. Experts from local agricultural bureaus and trained agronomy professionals will work together to label each image with pest/disease category, crop type, and stage of development.

A multi-stage review process will be implemented to ensure label accuracy. Additionally, we will incorporate a weak supervision mechanism: low-confidence or ambiguous images will first be automatically classified by a preliminary model, then manually verified. Selected hard cases will be escalated to expert review to determine whether to exclude them or assign proper labels.

5.3 Data Utilization and Feedback

After preprocessing and annotation, the data will be used to train a convolutional neural network (CNN) model based on MobileNetV3 Figure 9. The model will be continuously improved through semi-supervised learning driven by user feedback.

In cases where the model provides inaccurate predictions or encounters unfamiliar diseases not yet covered in the dataset, farmers can submit images directly via mobile upload. These samples will be reviewed by

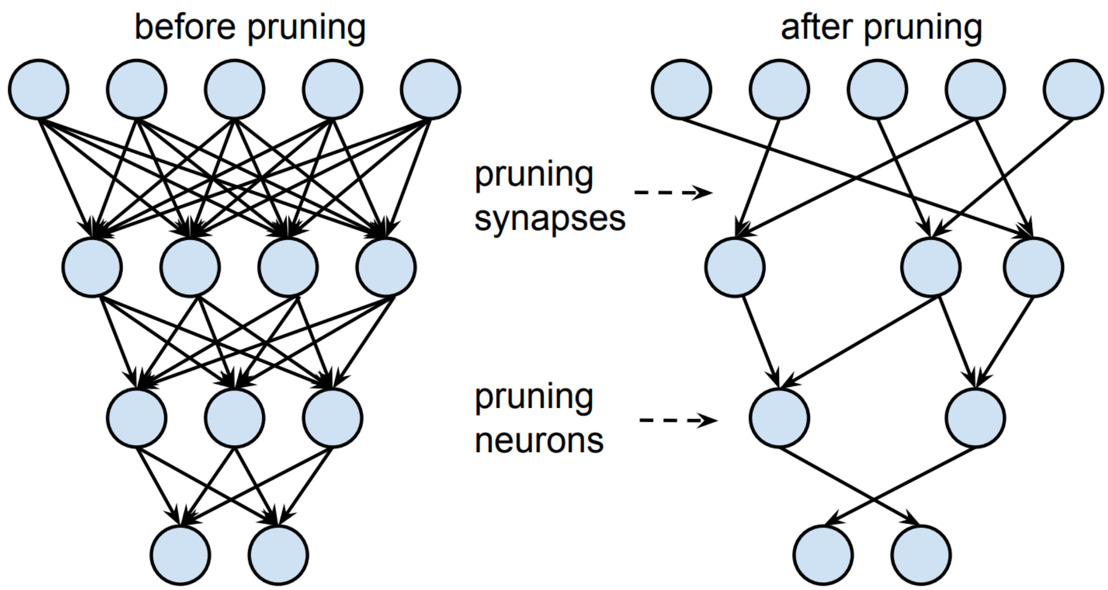

Figure 7: Pruning effects on neural networks.

an expert team, and validated cases will be added to the dataset and used to expand a retrieval-augmented generation (RAG) knowledge base for future reference and inference

6 Implementation Roadmap

Phase 1: Data Processing and Model Training (1–2 months)

In the first phase, we will follow the steps outlined in section 5 to preprocess and annotate the data, and proceed with model training. To meet the needs of individual farmers and address the specific requirements of pest and disease identification in crops, we will fine-tune a pre-trained MobileNetV3 model. Optimization techniques such as quantization and pruning will be applied to ensure the model can run efficiently on mobile CPUs, GPUs, or NPUs. Meanwhile, a cloud platform will be established in parallel, offering powerful computational resources to support enterprise-level demands.

Phase 2: Data Pipeline Construction (1–2 months)

We will build a complete data pipeline in this phase. Through a web interface, farmers will be able to easily upload images of misidentified or novel pests and diseases not covered by the existing database. These images will be automatically integrated into a local database in real time. A semi-automated annotation process will be implemented, combining weak supervision with expert review to ensure high labeling accuracy.

Phase 3: Pilot Testing (1 month)

This phase involves conducting a one-month pilot test. We will select a group of smallholder farmers for on-device testing and a group of enterprises for cloud AI testing. During the pilot, both groups will use the web and mobile apps to report incorrect predictions or newly identified pest/disease cases in real time. Leveraging RAG technology, the system will integrate local expert knowledge and historical data to further improve the accuracy and applicability of predictions.

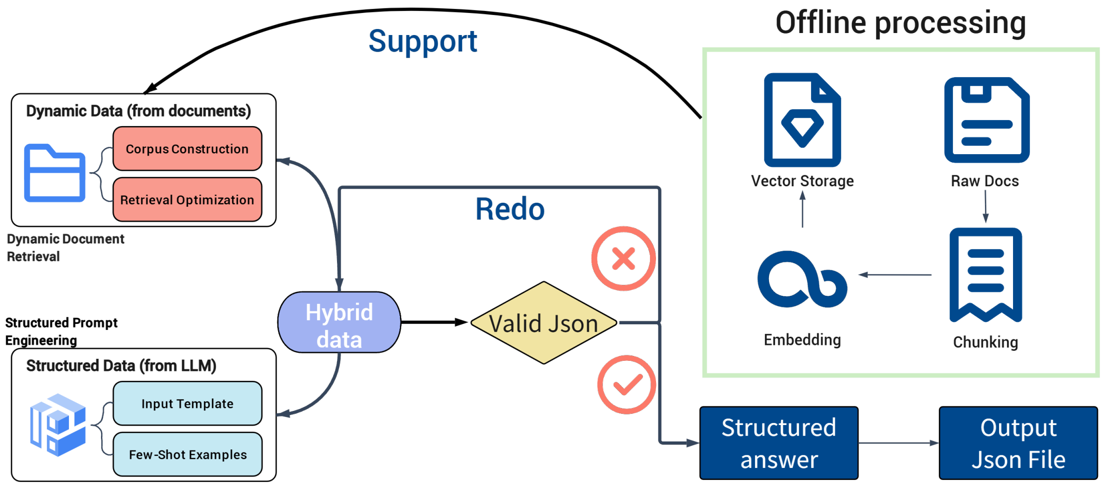

Figure 8: RAG workflow: vector search and generation.

Phase 4: Large-Scale Commercial Deployment (1 month)

Following a successful pilot, we will expand the on-device solution to broader rural regions, allowing users in areas with unstable internet access to continue using the mobile app for pest and disease identification and receive treatment suggestions. Enterprise users will also benefit from an enhanced cloud AI system capable of handling large-scale concurrent analyses and supporting multi-device collaboration.

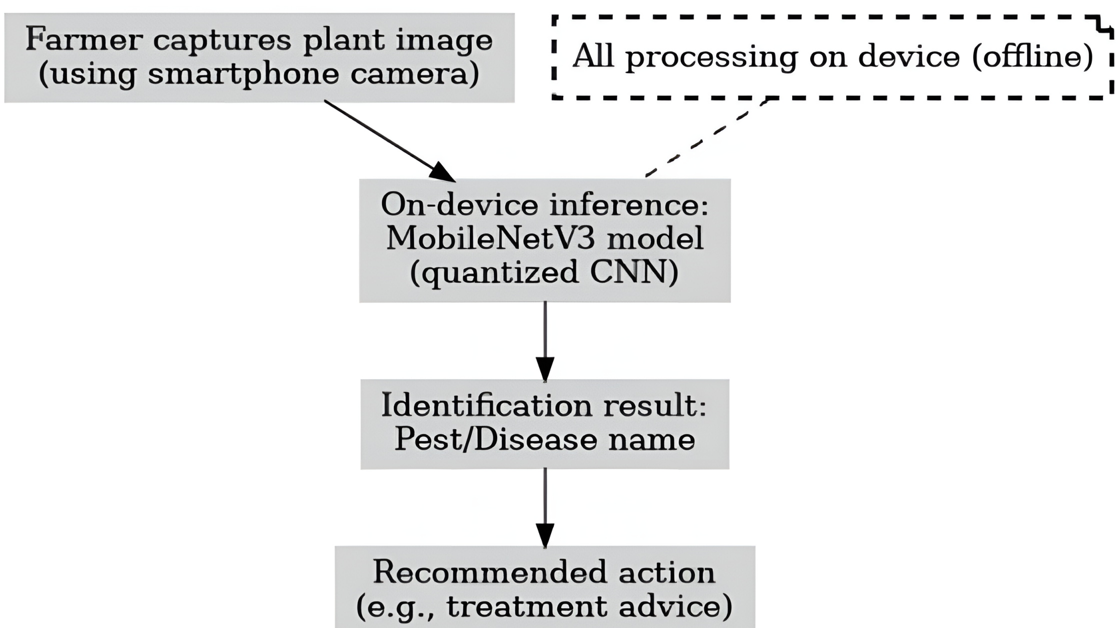

Figure 9: MobileNetV3 on-device inference flow for pest detection.

7 Ethical and Societal Impact

Deploying an AI solution in rural agriculture carries both opportunity and responsibility. While technology can empower farmers, it must be introduced in a way that respects their knowledge, privacy, and social context (Henrietta, 2023).

A primary concern is reliability. No AI model is perfect—misdiagnosis can lead to unnecessary pesticide use or, worse, untreated disease outbreaks. To mitigate this, our system is designed to be cautious: if the model's confidence is low, it simply says so, encouraging the farmer to consult an expert instead of offering a potentially misleading guess. Early versions of the app will also incorporate human-in-the-loop support, allowing difficult cases to be reviewed by extension officers when connectivity permits (Henrietta, 2023).

Another key consideration is data privacy. Farmers are rightfully cautious about where their field photos or usage patterns might go. That's why our model runs entirely on-device by default, with no need to send images to a server (Henrietta, 2023).

Crucially, we also recognize that not every farmer has equal access to digital tools. While smartphone adoption is widespread, it is not universal. To avoid widening the digital divide, we're working with local cooperatives to provide shared devices or app access through service points (Zhang et al., 2023).

We also take care to avoid unintended environmental or social harm. The app's recommendations follow integrated pest management principles, emphasizing non-chemical interventions where appropriate and avoiding advice that could lead to excessive pesticide use (Albanese et al., 2021). All guidance is vetted to align with agricultural regulations and best practices, regardless of any external partnerships.

Ultimately, the Smart Agricultural System is not about replacing traditional expertise—it's about extending it. By respecting local knowledge, enabling informed decisions, and staying transparent about limitations, we aim to build a tool that complements human judgment and builds trust in both the technology and those who use it (Yao et al., 2023).

Code Availability

All code and evaluation scripts used in this work are publicly available at: `https://marekkel.github.io/MIS3011_Project/`.

References

[1] Donnellon-May, G. (2025, February 27). *China's grand plan for food self-sufficiency*. Asia Times. `https://asiatimes.com/2025/02/chinas-grand-plan-for-food-self-sufficiency/`

[2] FAO. (2020). *International Year of Plant Health brochure*. Food and Agriculture Organization of the United Nations. `https://www.fao.org/plant-health-2020/home/en/`

[3] Wang, W. (2023, March 3). *Unlocking the potential of smallholder farmers*. China Daily Global. `https://www.chinadaily.com.cn/`

[4] Zhang, P., Wang, J., & Li, X. (2023). Implementation of Artificial Intelligence in Agriculture. *Journal of Computational and Cognitive Engineering*, 2(2), 155–162.

[5] Henrietta, H. M. (2023). Artificial Intelligence in Agriculture: A Review of Current Applications and Future Trends. *Journal of Computational and Cognitive Engineering*, 2(2), 154–165.

[6] Albanese, A., Nardello, M., & Brunelli, D. (2021). Automated pest detection with DNN on the edge for precision agriculture. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 11(3), 431–440.

[7] Yao, J., Tran, S. N., Sawyer, S., & Garg, S. (2023). Machine learning for leaf disease classification: data, techniques and applications. *Artificial Intelligence Review*, 56(8), 3571–3616.

[8] Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., & Shen, Y. (2015). Managing nitrogen for sustainable development. *Nature*, 528(7580), 51–59.

[9] Tian, X. (2020). Current status and recent trend of irrigation water use in China. *International Commission on Irrigation and Drainage.*

[10] Li, Y., Wang, Y., & Liu, X. (2021). The impact of rural population aging on farmers' cleaner production behavior. *Sustainability*, 13(15), 8532.

[11] Yicai Global. (2022, September 23). Nearly a quarter of China's elderly are still working, most as farmers.

[12] Pereira, L.S., et al. (2021). Water Productivity Gains Through AI-Driven Irrigation: A Global Meta-Analysis. *Agricultural Water Management*, 256, 107081.

[13] Grafton, R.Q., et al. (2018). The Paradox of Water Pricing: Unbundling Water Rights Through Blockchain. *Nature Sustainability*, 1(9), 466–473.

[14] van Lenteren, J.C., et al. (2020). AI-Guided Predator-Prey Dynamics Monitoring in Greenhouse Ecosystems. *Biosystems Engineering*, 198, 1–12.

[15] Carter, M.R., et al. (2021). Index-Based Insurance with Machine Learning: Evidence from East Africa. *American Journal of Agricultural Economics*, 103(5), 1627–1649.

[16] Dias, A., et al. (2022). AI-Driven Crop Switching Reduces Income Volatility: Evidence from Brazil. *Food Policy*, 112, 102365.

[17] International Coffee Organization (ICO) (2023). Annual Review: Price Transmission Mechanisms in Coffee Markets.

[18] Kozai, T., et al. (2022). Beyond Photosynthetic Limits: AI-Optimized LED Spectra in Plant Factories. *Plant Biotechnology Journal*, 20(3), 511–525.

[19] USDA ERS. (2021). Price Premiums for AI-Certified Crops: A Market Analysis.

[20] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3.

[21] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70–90.

[22] Plantix. (2023). How Plantix works. PEAT GmbH. `https://plantix.net/en/how-it-works/`

[23] World Economic Forum. (2021). The Future of Agriculture: AI and Robotics. `https://www.weforum.org`

[24] McKinsey & Company. (2022). Agriculture's digital revolution: Why AI is the next frontier. `https://www.mckinsey.com/industries/agriculture/our-insights`